

# Exploring Probability

Patch Kessler, June 14, 2008

## 1 Fundamentals

### 1.1 Sets, Lists, and Counting

The set  $A$  is equal to the set  $B$  iff  $x \in A \iff x \in B$ . From this it follows that  $\{a\} = \{a, a\}$ , and so to avoid confusion, we say that a set “contains  $n$  items” iff it contains  $n$  distinct items. A list is a finite ordered set; note that  $(a) \neq (a, a)$ .

Let  $l(n, k)$  denote the number of ways to construct a list of  $k$  distinct items from a set of  $n$  items. There are  $n$  ways to choose the first list element,  $n - 1$  ways to choose the second element,  $n - 2$  ways to choose the third, and so on, with  $n - (k - 1)$  ways to choose the  $k^{\text{th}}$  element. It follows that

$$l(n, k) = n \cdot (n - 1) \cdot (n - 2) \cdots (n - k + 1) = \frac{n!}{(n - k)!}. \quad (1)$$

When  $k = n$ , the left equality shows that  $l(n, n) = n!$ , and then the right equality shows that we should adopt the convention  $0! = 1$ .

Let  $\binom{n}{k}$  (spoken as “ $n$  choose  $k$ ”) denote the number of ways to construct a set of  $k$  items from a set of  $n$  items. Each of these sets of  $k$  items can be used to construct  $k!$  lists of  $k$  items, and so the total number  $l(n, k)$  of lists of  $k$  items is given by

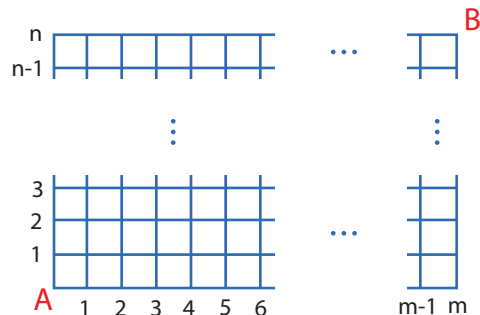
$$l(n, k) = k! \binom{n}{k}. \quad (2)$$

Together with (1), this allows us to conclude that

$$\binom{n}{k} = \frac{n!}{k!(n - k)!} = \frac{n \cdot (n - 1) \cdot (n - 2) \cdots (n - k + 1)}{1 \cdot 2 \cdot 3 \cdots k}. \quad (3)$$

### 1.2 Paths Through a Grid

Let  $p(m, n)$  be the number of paths from  $A$  to  $B$  in the grid below, moving only up and to the right.



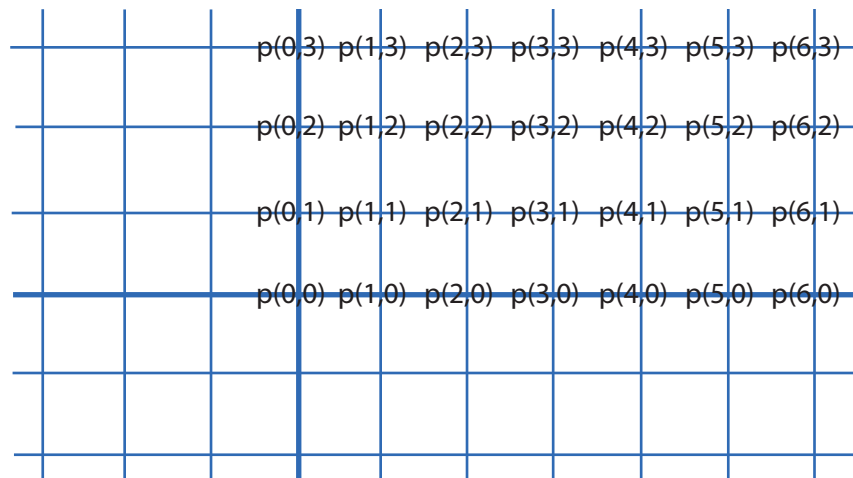
We think of each path from  $A$  to  $B$  as the deformation of a physical chain of  $m + n$  line segments. The subset of chain segments which are horizontal completely describes the chain’s path, and so the number of paths  $p(m, n)$  equals the number of possible subsets, which is  $\binom{m + n}{m}$  (the total number of chain segments) choose  $m$  (the number of segments which are horizontal). From 1.1, this quantity is given by

$$p(m, n) = \binom{m + n}{m} = \frac{(m + n)!}{m! n!}. \quad (4)$$

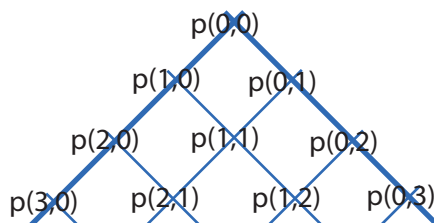
The argument is symmetric in  $m$  and  $n$ , and so it is no surprise that  $p(m, n) = p(n, m)$ .

### 1.3 Pascal's Triangle

Assign  $p(m, n)$  to the node located by  $(m, n)$  in the cartesian plane, where  $m$  and  $n$  are non-negative integers.



This array of numbers is called *Pascal's Triangle*, and is usually rotated  $135^\circ$  as follows:



From (3), it follows that  $\binom{n}{k} = \frac{n}{n-k} \binom{n-1}{k}$ . Together with (4), this allows us to write

$$\begin{aligned} p(m, n) &= \frac{m+n}{m} \binom{m+n-1}{n} = \frac{m+n}{m} p(m-1, n), \\ p(m, n) &= \frac{m+n}{n} \binom{m+n-1}{m} = \frac{m+n}{n} p(m, n-1), \end{aligned} \quad (5)$$

from which it follows that

$$p(m, n) = p(m-1, n) + p(m, n-1). \quad (6)$$

The numbers in Pascal's triangle therefore satisfy a discrete 2D boundary value problem. Every value can be generated using (6) together with the boundary conditions

$$p(0, 0) = 1, \text{ and } p(m, n) = 0 \text{ when } m < 0 \text{ or } n < 0. \quad (7)$$

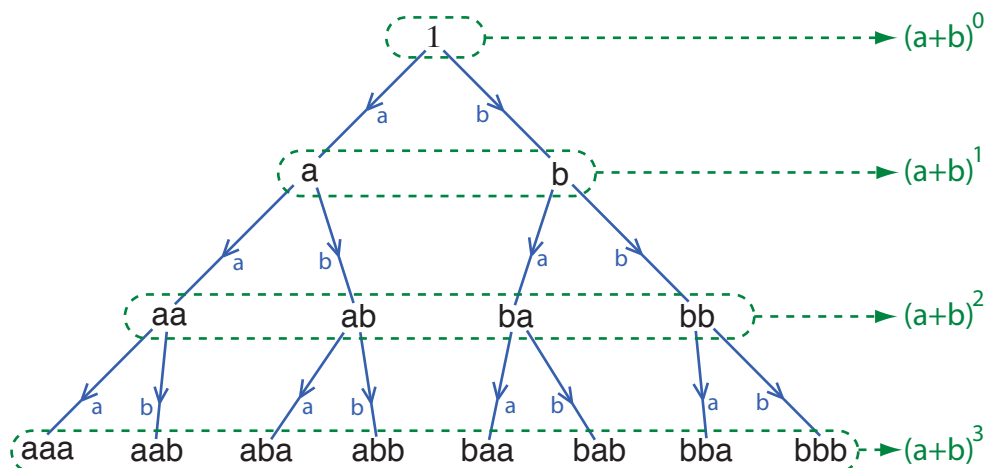
This is the simplest recipe for Pascal's Triangle, however it makes it awkward to move from (6) to (4). *Defining* the triangle in terms of possible paths through a grid allows for a more direct development of formulas, and also highlights a way to see the triangle that will be useful in later sections.

## 1.4 The Binomial Theorem

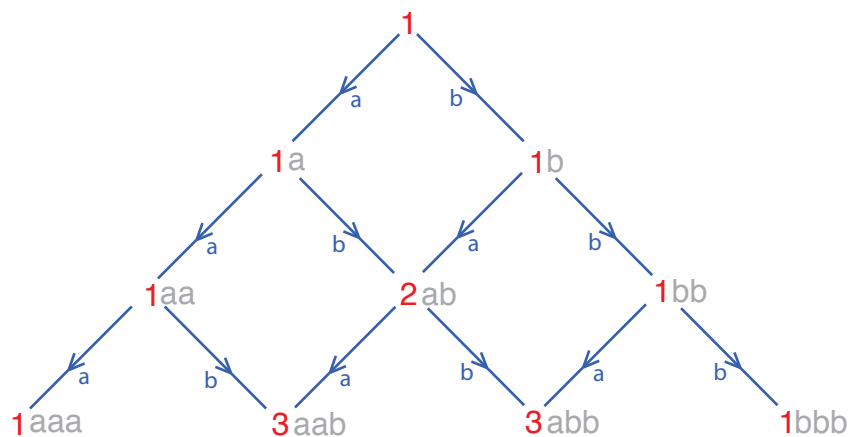
A notation is a convention for conveying information by positioning symbols in two-dimensional space. The notation  $(a + b)^N$  tells us to

*add the quantities  $a$  and  $b$ , and then multiply the result by itself  $N$  times.*

We obtain a deeper understanding of the terms that arise in an *expansion* of  $(a + b)^N$  by switching to a different notation, in which arrows act as product operators. In this new notation, we associate the product  $xa$  with the tip of an arrow (labeled  $a$ ) which points away from  $x$ , (e.g.,  $x \xrightarrow{a} xa$ ). In the case of repeated multiplications of  $(a + b)$  this notation gives rise to a cascading tree of symbols



Let  $N$  denote the tree row number, with  $N = 0$  at the tree apex. Starting with 1 at the apex, we obtain row  $N$  by multiplying each term in row  $N - 1$  by  $(a + b)$ . The terms in row  $N$  therefor sum to  $(a + b)^N$ , as noted in green. We put the tree in Pascal Triangle form by consolidating node cousins row by row as the tree is constructed. (Node cousins are nodes with the same grandparent but different parents, such as  $ab$  and  $ba$ .)



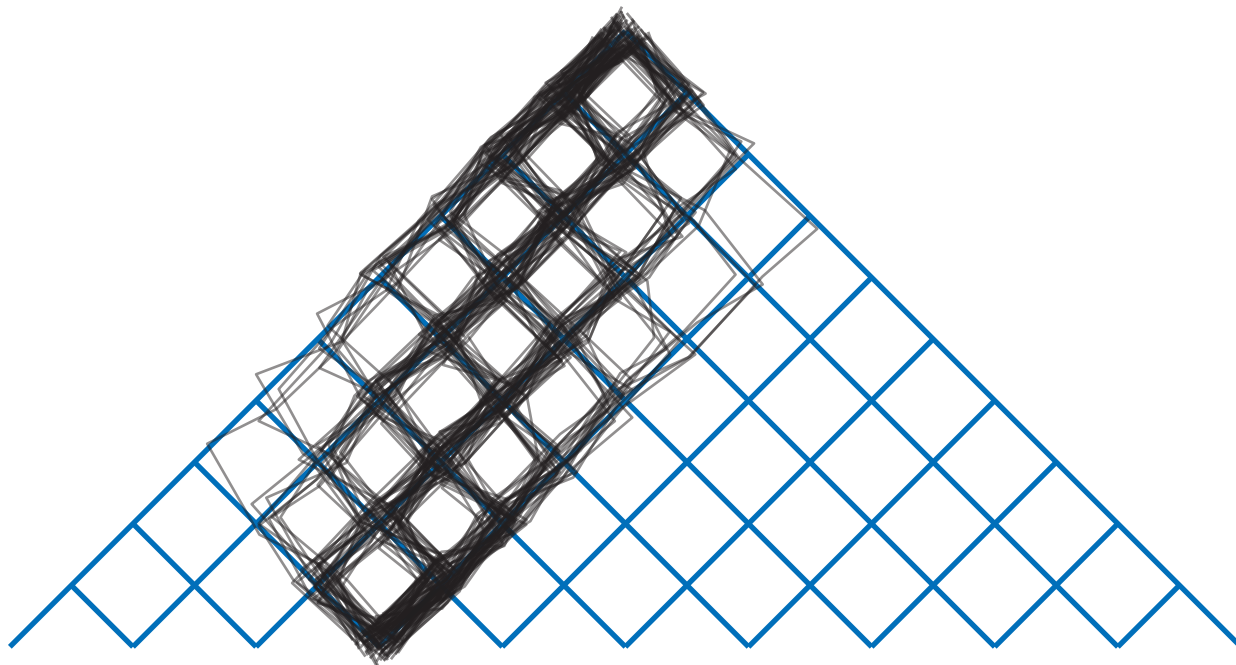
In row  $N$  of the consolidated tree, the  $k^{\text{th}}$  node from the left (starting at 0) is given by  $B(N, k)a^{N-k}b^k$ , where  $B(N, k) = B(N - 1, k) + B(N - 1, k - 1)$  within the tree, and  $B = 1$  along the tree sides (i.e., whenever  $k$  equals 0 or  $N$ ). Switching to coordinates  $m = N - k$  and  $n = k$ , we see that these are exactly the conditions (6) and (7) that generate Pascal's Triangle. An explicit formula for  $B(N, k)$  is therefor given by (4)

$$B(N, k) = p(N - k, k) = p(k, N - k) = \binom{N}{k}. \quad (8)$$

The terms in row  $N$  sum to  $(a + b)^N$ , and so

$$(a + b)^N = \binom{N}{0} a^N b^0 + \binom{N}{1} a^{N-1} b^1 + \binom{N}{2} a^{N-2} b^2 + \dots + \binom{N}{N} a^0 b^N. \quad (9)$$

This result is called the Binomial Theorem. The coefficient  $\binom{N}{k}$  is the number of possible paths from the tree apex to the  $k^{\text{th}}$  node on row  $N$ . As an example, the  $\binom{10}{3} = 120$  distinct paths from the tree apex to the node located by  $N = 10$  and  $k = 3$  are shown below. This number is small for nodes near the row ends (i.e.,  $k$  close to 0 or  $N$ ), and is big for nodes near the row center.



## 2 Probability and Expectation

Consider a broken copy machine spitting out pages that are either totally white or totally covered in ink (i.e., black). Suppose that the probability of getting a white page is  $a \in [0, 1]$ , and that the probability of getting a black page is  $b = 1 - a$ . This kind of statement is encountered so frequently that it's tempting to accept it without thinking, however here we pause to ponder it. Informally, the closer  $a$  is to 1, the more we *expect* to see a white page. However if all we ever see is a few pages from the machine, then the notion of expectation breaks down. For instance if we only get one page from the machine, then what does it matter what we expected the color to be? Suppose we expected black, and the single page is white. Does greater expectation mean greater surprise at a contrary result, or a greater sense of "I knew that's what it was going to be!" at an expected result? If we could put surprise into correspondence with the real numbers from 0 to 1 then it would make sense to quantify expectation for a single trial. I submit that we can't, and that probability only makes sense in the context of repeated trials (either real or imagined).

To repeat a trial is to open another can of worms. If the copy machine was to do exactly the same thing a second time, it would output the same color paper. We want the machine to *behave* the same but for the second trial to be different. We make this rigorous in a later section.

### 2.1 Bernoulli Trials

Lots of processes exhibit the branching structure of Pascal's triangle.

## 2.2 The case of big $N$

### 3 Axiomatic Probability

A *probability space* consists of three items,

- a set  $\Omega$  of *sample points*.
- a  $\sigma$ -algebra  $U$  of subsets (called *events*) of  $\Omega$ .
- a mapping  $P : U \rightarrow [0, 1] \subset \mathbb{R}$  with certain properties which we call a *measure* on  $U$ .

### 4 Independence

A collection of events is called *independent* if the probability that a sample point is contained in one event isn't affected by knowing that the sample point is in any selection of the remaining events. Calling the one event  $A$ , and the intersection of the selected remaining events  $B$ , we need

$$P(A) = \frac{P(A \cap B)}{P(B)}. \quad (10)$$

This equation has to hold for all possible choices of  $A$ , and for all possible choices of the events that intersect to form  $B$ . Suppose the total number of events in the collection is  $n$ . If  $B$  is the intersection of  $k$  events, then there are  $n \binom{n-1}{k}$  ways to write out (10). The total number of explicit equations needed to capture independence is therefore given by

$$n \binom{n-1}{1} + n \binom{n-1}{2} + \cdots + n \binom{n-1}{n-1} = n2^{n-1} - n \quad (11)$$

This set of conditions is REDUNDANT!!! (but clear)

#### 4.1 Alternate Forms

For the collection of events to be independent, (10) has to hold for *all* choices of  $A$  and  $B$ . Including the  $n$  trivial equations in which  $B = \emptyset$ , the number of equations needed to describe independence is  $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n$ . The  $k = 1$  equation  $P(A_{i_1} \cap A_{i_2}) = P(A_{i_1})P(A_{i_2})$  can be combined with the  $k = 2$  equation  $P(A_{i_1})P(A_{i_2} \cap A_{i_3}) = P(A_{i_1} \cap A_{i_2} \cap A_{i_3})$  to obtain

$$P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) = P(A_{i_1})P(A_{i_2})P(A_{i_3}). \quad (12)$$

Continuing in this way, we can express independence with  $n - 1$  (nontrivial) sets of equations, each of the form

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}) \quad (13)$$

for  $k$  from 2 to  $n$ , where  $\{i_1, \dots, i_k\}$  stands for each of the  $\binom{n}{k}$  sets of  $k$  integers that can be chosen from  $1 : n$ .